

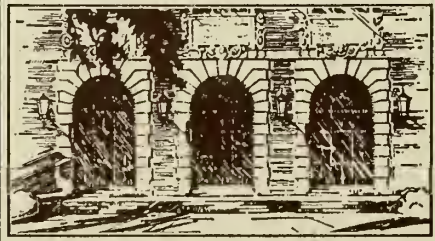
LIBRARY OF THE
UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

510.84

I l 6 r

no. 160-170

cop. **2**




The person charging this material is responsible for its return to the library from which it was withdrawn on or before the **Latest Date** stamped below.

Theft, mutilation, and underlining of books are reasons for disciplinary action and may result in dismissal from the University.

UNIVERSITY OF ILLINOIS LIBRARY AT URBANA-CHAMPAIGN

NOV 15 1973

NOV 15 1973



Digitized by the Internet Archive
in 2013

<http://archive.org/details/hybridmethodsfor164gear>

510.84
I 16r
no. 164
cop. 3

286-415-1012

UNIVERSITY OF ILLINOIS
GRADUATE COLLEGE
DIGITAL COMPUTER LABORATORY

REPORT NO. 164

HYBRID METHODS FOR INITIAL VALUE PROBLEMS IN ORDINARY
DIFFERENTIAL EQUATIONS

by
C. W. Gear

June 19, 1964

no. 107
cap.
Abstract

Methods for the integration of initial value problems for the ordinary differential equation $\frac{dy}{dx} = f(x,y)$ which are a combination of one step procedures (eg., Runge-Kutta) and multistep procedures (eg., Adams' Method) are discussed. A generalization of a theorem from Henrici ⁽³⁾ proves that these methods converge under suitable conditions of stability and consistency. This, incidentally, is also a proof that predictor-corrector methods using a finite number of iterations converge. Four specific methods of order 4,6,8 and 10 have been found. Numerical comparisons of the first three of these have been made with Adams', Nordsieck's and the Runge-Kutta methods.

1. Introduction

The original motivation for this work was a desire to achieve some of the flexibility of Runge-Kutta type methods, which allow easy starting and step size changing procedures, and at the same time achieve the increased speed due to the fewer function evaluations and the higher order possible with multistep methods such as Adams' method. Multistep predictor-corrector methods are also characterized by readily yielding a number (the difference between the predictor and the corrector) that is a reasonable estimate of the local truncation error. This number can be used to automatically control the step size. Runge-Kutta methods typically require a further function evaluation to get such an estimate. [e.g., see the Kutta-Merson process, Fox⁽¹⁾ page 24].

One approach to this problem has been made by Nordsieck⁽⁴⁾ who recasts Adams' method in such a way that the derivatives of the approximating polynomial, rather than information at several function points, are retained. These are independent of step size so that it can be changed at will. The original approach made by this author was the following.

- 1) Reduce the amount of information (the number of points) being retained to minimize the start-up and step change problems.
- 2) Calculate information at the mid points as part of the process. Doubling the step size requires no interpolation. If the mid-points are available, halving would also require no interpolation (although it is true that interpolation requires very little arithmetic in comparison to the rest of the job).

Therefore, the investigation started with the class of methods described by

$$y_{n+\frac{1}{2}} = \sum_{i=0}^k \left[\alpha_{0i} y_{n-i} + h \beta_{0i} f_{n-i} \right]$$

$$f_{n+\frac{1}{2}} = f\left(x_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}\right)$$

$$y_{n+1}^p = \sum_{i=0}^k \left[\alpha_{1i} y_{n-i} + h\beta_{1i} f_{n-i} \right] + \gamma_{10} h f_{n+\frac{1}{2}}$$

$$f_{n+1} = f(x_{n+1}, y_{n+1}^p)$$

$$y_{n+1} = \sum_{i=0}^k \left[\alpha_{2i} y_{n-i} + h\beta_{2i} f_{n-i} \right] + h \left[\gamma_{20} f_{n+\frac{1}{2}} + \gamma_{21} f_{n+1} \right] \quad (1.1)$$

applied to the differential equation

$$\frac{dy}{dx} = f(x, y)$$

The derivative at x_{n+1} is not recalculated from the corrected value of y_{n+1} in order to save a function evaluation.

These methods require knowledge about k points in addition to x_n . They have been investigated in detail for $k = 1$ and 2 and partially for k up to 9 . The main numerical result is that it is possible to achieve a high order accuracy of $(2k + 2)$ for k up to 4 , but not for k from 5 to 9 . Since this order accuracy compares favorably with the $k + 2$ ($k + 3$ if k odd) accuracy obtainable from multistep methods, the emphasis was shifted from the problem of making life easier for changing the step size to that of achieving high order methods using few points.

Four specific methods of the class (1.1) are given. One has a fourth order corrector formula with $k = 1$. It has one free parameter in the corrector which must be chosen to make the method stable. In Section 3, the extraneous roots of the linear difference equations for the error are discussed. There are $2k + 1$ non-principal roots, of which $k + 1$ are zero if $h \frac{\partial f}{\partial y} = 0$. For $k = 1$ the other non-principal root can also be made equal to zero by a suitable choice of the parameter.

When $k = 2$, a 6th order method can be found. It also has one free parameter in the corrector which can be chosen to minimize the extraneous roots of the difference equation. Zero roots can only be achieved by returning to a 5th order method. The predictor y_{n+1}^p also has one free parameter if it is 5th order (which will only cause a 7th order error in the corrector). This parameter can be chosen to get 6th order in the predictor, but it then has

undesirable stability effects on the method. In Section 4, various choices of these parameters are discussed, and the method is compared with the Adams 6th order method, Nordsieck's method and the Runge-Kutta method on the integration of $J_{16}(x)$ for $x = 6(h) 6138$, with $h = 1/32, 1/16$ and $1/8$.

Section 5 discusses higher order methods that can be achieved. Coefficients of 8th and 10th order methods which are stable are given. The 8th order method was also used to integrate $J_{16}(x)$, $x = 6(1/16) 6138$ with good accuracy. The 10th order method was not used as it almost certainly has undesirable stability properties when $\frac{\partial f}{\partial y}$ is non-zero.

A recent paper by Gragg and Steller⁽²⁾ deals with a generalization of this method where the non mesh-point may be any point fixed in relation to x_n . The generalization taken in Section 2 of this paper allows a number of non mesh-points to be used. A theorem is proved giving sufficient conditions for the convergence of these methods. These conditions are that each of the predictors (the estimates of the solution at a set of points) is at least of order zero, that the corrector (the final estimate of y and the next mesh point) is at least of order 1 and that the corrector is stable. This general formulation contains Runge-Kutta and multistep predictor-corrector methods as subsets. The formulation is explicit since, in practice, only a finite number of iterations of the corrector can be used. All but the last can be viewed as a set of predictors.

2. A General Formulation and Proof of Convergence

Consider the sequence of operations

$$p_{0,n+1} = \sum_{i=0}^k \left[\alpha_{0i} y_{n-i} + h\beta_{0i} f_{n-i} \right]$$

and

$$p_{j,n+1} = \sum_{i=0}^k \left[\alpha_{ji} y_{n-i} + h\beta_{ji} f_{n-i} \right] + h \sum_{i=0}^{j-1} \gamma_{ji} f_{n-i} \quad (2.1)$$

for

$$j = 1, 2, \dots, J$$

where h is the step size ($x_n = a + nh$) and $f_{p_i} = f(p_{i,n+1})$ (the quantities y , f and p are assumed to be vectors representing both the independent variable x and the one or more components of the dependent variable). $p_{J-1,n+1}$ will be taken as the predicted value of y_{n+1} and $p_{J,n+1}$ will be taken as the corrected value. To avoid a final evaluation of the derivative f , we define f_{n+1} by $f_{n+1} = f(p_{J-1,n+1})$.

If $k = 0$, this method is the general explicit Runge-Kutta method. On the other hand, if the coefficients are such that

$$\left. \begin{aligned} \alpha_{j,i} &= \alpha_{j-1,i} \\ \beta_{j,i} &= \beta_{j-1,i} \\ \gamma_{j,j-1} &= \gamma_{j-1,j-2} \end{aligned} \right\} \begin{aligned} &\text{for } i = 0, 1, \dots, k \\ &\text{and } j = 2, 3, \dots, J \end{aligned}$$

and all other γ_{ij} are zero then this method represents the use of a multistep predictor followed by J applications of a corrector formula.

To discuss the error and convergence of this method we introduce the usual notation $\epsilon_n = y_n - y(x_n)$ where $y(x)$ is the solution of the differential equation, and y_n is the value at $x = x_n$ calculated by equations (2.1) from a suitable set of starting values y_0, y_1, \dots, y_k . We use the definition of convergence from Henrici⁽³⁾, that is, the method is convergent if, for every problem $\frac{dy}{dx} = f(x, y)$, $x \in [a, b]$, $y(a) = A$, satisfying Lipschitz and continuity conditions for $x \in [a, b]$ and $y \in (-\infty + \infty)$, $\epsilon_n \rightarrow 0$ as $x_n \rightarrow x \in [a, b]$ and $y_i \rightarrow A$ for $i = 0, \dots, k$ as $h \rightarrow 0$.

The method is said to be stable if each of the roots of the polynomial equation $\rho(\xi) = \xi^{k+1} - \sum_{i=0}^k \xi^{k-i} \alpha_{ji} = 0$ is either inside the unit circle or on the unit circle and simple.

The method is consistent if the corrector is of order 1 or greater and if all of the predictors, $p_{i,n+1}$, $i = 0, \dots, J-1$, are of order zero or greater. Formally this means that

$$1 - \sum_{i=0}^k \alpha_{ji} = 0 \quad (2.2)$$

for $j = J-1, J$ and for those j such that $\gamma_{Jj} \neq 0$ (those $p_{j,n+1}$ that are not used in the corrector do not even have to be of order 0 and if the $\beta_{J,i}, i = 0, \dots, k$ and $\gamma_{J,J-1}$ are 0, (2.2a) need not hold for $j = J-1$)

$$\text{and} \quad k+1 - \sum_{i=0}^k (k-i) \alpha_{Ji} = \sum_{i=0}^k \beta_i + \sum_{i=0}^{J-1} \gamma_{Ji}. \quad (2.2b)$$

The theorem to be proved is that "the Method Converges if it is Stable and Consistent." The proof closely parallels the proof of a similar theorem in Henrici⁽³⁾ (Theorem 5.10).

Proof. Define

$$p_{j,n+1}^T = \sum_{i=0}^k \left(\alpha_{ji} y(x_{n-i}) + h \beta_{ji} f(y(x_{n-i})) \right) + h \sum_{i=0}^{j-1} \gamma_{ji} f(p_{i,n+1}^T) \quad (2.3)$$

for

$$j = 0, 1, \dots, J.$$

Let

$$\left. \begin{aligned} L_h^p(x_n, y) &= p_{J-1,n+1}^T - y(x_{n+1}) \\ L_h(x_n, y) &= p_{J,n+1}^T - y(x_{n+1}) \end{aligned} \right\} \quad (2.4)$$

and

Thus L_h^p and L_h are the truncation errors in the final predictor and corrector respectively. We will first show that consistency implies the L_h^p and $L_h/h \rightarrow 0$ as $h \rightarrow 0$ uniformly for $x \in [a, b]$ and then shown that these conditions and stability imply convergence.

Define

$$X(\delta) = \max_{\substack{|x-x^*| \leq \delta \\ x, x^* \in [a, b]}} |y'(x^*) - y'(x)|$$

$X(\delta)$ exists and $\rightarrow 0$ as $\delta \rightarrow 0$ since $\frac{dy}{dx} = f(x, y(x))$ exists and is continuous in the closed interval $[a, b]$.

Hence

$$y'(x_{n-i}) = y'(x_n) + \theta_i X(ih)$$

and

$$y(x_{n-i}) = y(x_n) - ih \left[y'(x_n) + \theta'_i X(ih) \right]$$

where

$$|\theta_i| \leq 1, \quad |\theta'_i| \leq 1 \text{ and } x_n, x_{n-i} \in [a, b].$$

Now

$$\begin{aligned} L_h^T p(x_n, y) &= p_{J-1, n+1}^T - y(x_{n+1}) \\ &= \sum_{i=0}^k \alpha_{J-1, i} [y(x_n) - ih[y'(x_n) + \theta'_i X(ih)]] \\ &\quad + h \sum_{i=0}^k \beta_{J-1, i} [y'(x_n) + \theta_i X(ih)] \\ &\quad + h \sum_{i=0}^{J-2} \gamma_{J-1, i} f(p_i^T, n+1) \\ &\quad - y(x_n) - h[y'(x_n) + \theta'_{-1} X(h)] \end{aligned}$$

But

$$\sum_{i=0}^k \alpha_{J-1, i} = 1$$

Therefore

$$L_h^p(x_n, y) = O(h) \rightarrow 0 \text{ as } h \rightarrow 0$$

Similiary

$$\begin{aligned} L_h(x_n, y) &= p_{J, n+1}^T - y(x_{n+1}) \\ &= y(x_n) \left[\sum_{i=0}^k \alpha_{J,i} - 1 \right] + h \left[-y'(x_n) - \sum_{i=0}^k i \alpha_{J,i} y'(x_n) + \sum_{i=0}^k \beta_{J,i} y'(x_n) \right. \\ &\quad \left. + \sum_{i=0}^{J-1} \gamma_{J,i} f\left(\sum_{m=0}^k \alpha_{i,m} y(x_n) + O(h)\right) \right] + hX(kh) B \end{aligned}$$

where B is bounded as $h \rightarrow 0$.

If $\gamma_{J,i} \neq 0$, then $\sum_{m=0}^k \alpha_{i,m} = 1$, and f satisfies a Lipshitz condition, so that the last term can be replaced by

$$\sum_{i=0}^{J-1} \gamma_{J,i} y'(x_n) + O(h)$$

Now using

$$1 = \sum_{i=0}^k \alpha_{J,i}$$

and

$$1 + \sum i \alpha_{J,i} - \sum \beta_{J,i} - \sum \gamma_{J,i} = k + 1 - \sum (k-i) \alpha_{J,i} - \sum \beta_{J,i} - \sum \gamma_{J,i} - k(1 - \sum \alpha_{J,i}) = 0$$

we get

$$L_h(x_n, y) = hBX(kh) + O(h^2)$$

$$\therefore L_h/h \rightarrow 0 \text{ as } h \rightarrow 0$$

Define the error in the predictor as ϵ_n^* , that is, $\epsilon_n^* = p_{J-1,n} - y(x_n)$.

Then

$$\epsilon_{n+1} = p_{J,n+1} - p_{J,n+1}^T + L_h(x_n, y) \quad (2.5)$$

and

$$\epsilon_{n+1}^* = p_{J-1,n+1} - p_{J-1,n+1}^T + L_h^p(x_n, y) \quad (2.6)$$

When we substitute for the $p-p^T$ terms in (2.5) and (2.6), terms of the form $h\gamma(f(p_{J,i}) - f(p_{J,i}^T))$ will be included, which, by the Lipschitz condition, can be replaced by $h\gamma L(p_{J,i} - p_{J,i}^T)$ where L is bounded. Since the formula is explicit, the $p_{J,i} - p_{J,i}^T$ terms can be substituted for repeatedly. The process stops after a maximum of J steps, and results in a polynomial in h with bounded coefficients.

Thus (2.5) and (2.6) can be rewritten as

$$\epsilon_{n+1} = L_h(x_n, y) + \sum_{i=0}^k \alpha_{J,i} \epsilon_{n-i} + h \sum_{i=0}^k (\epsilon_{n-i} P_{1i}(h, x_n) + \epsilon_{n-i}^* P_{2i}(h, x_n)) \quad (2.7)$$

and

$$\epsilon_{n+1}^* = L_h^p(x_n, y) + \sum_{i=0}^k \alpha_{J-1,i} \epsilon_{n-i} + h \sum_{i=0}^k (\epsilon_{n-i} P_{3i}(h, x_n) + \epsilon_{n-i}^* P_{4i}(h, x_n)) \quad (2.8)$$

where the P_{1i} , P_{2i} , P_{3i} and P_{4i} are polynomials in h with bounded coefficients.

Henrici⁽³⁾, lemma (5.5) shows that if

$$\xi^{k+1} - \sum_{i=0}^k \alpha_{J,i} \xi^{k-i}$$

is the polynomial of a stable method, and if

$$\frac{1}{1 - \sum_{i=0}^k \alpha_{J,i} \xi^{i+1}} = \sum_{p=0}^{\infty} \partial_p \xi^p \quad (2.9)$$

then \exists a constant $\Gamma < \infty$ such that $|\partial_p| \leq \Gamma, p = 0, 1, \dots$. Multiply equation (2.7) by ∂_{N-n-1} and sum for $n = k$ to $N-1$ to get

$$\begin{aligned} & \epsilon_N \delta_0 + \epsilon_{N-1} (\delta_1 - \alpha_{J,0} \delta_0) + \epsilon_{N-2} (\delta_2 - \alpha_{J,0} \delta_1 - \alpha_{J,1} \delta_0) + \dots \\ & + \epsilon_{k+1} (\delta_{N-k-1} - \alpha_{J,0} \delta_{N-k-2} - \dots - \alpha_{J,k} \delta_{N-2k-2}) \\ & = \sum_{n=k}^{N-1} \delta_{N-n-1} L_h(x_n, h) + \sum_{i=0}^k (\text{Bounded multiples of } \epsilon_i \text{ and } \epsilon_i^*). \\ & + h \sum_{n=k}^{N+1} \delta_{N-n-1} \sum_{i=0}^k \left[e_{n-i} P_{1i}(h, x_n) + \epsilon_{n-i}^* P_{2i}(h, x_n) \right] \end{aligned} \quad (2.10)$$

From (2.9), $\delta_0 = 1$, and terms like

$$\delta_m - \alpha_{J,0} \delta_{m-1} - \alpha_{J,1} \delta_{m-2} - \dots - \alpha_{J,k} \delta_{m-k-1} = 0$$

Therefore, the left hand side of (2.10) = ϵ_N . The last term of the right hand side involves each ϵ_i and ϵ_i^* no more than $k+1$ times. Therefore (2.10) gives the bound

$$|\epsilon_N| \leq \Gamma(N - k) \max |L_h(x_n, y)| + M \sum_{i=0}^k (|\epsilon_i| + |\epsilon_i^*|) + hC \sum_{n=0}^{N-1} (|\epsilon_n| + |\epsilon_n^*|)$$

where M and C are bounded for all $h \leq$ some h_0 . Now $N-k \leq \frac{b-a}{h}$, and

$$|L_h(x_n, y)/h| \leq \mu(h) \rightarrow 0 \text{ as } h \rightarrow 0,$$

$$\therefore |\epsilon_N| \leq (b-a)\Gamma\mu(h) + S(h) + hC \sum_{n=0}^{N-1} (|\epsilon_n| + |\epsilon_n^*|) \quad (2.11)$$

here $S(h)$ is a function of the error in the initial conditions which $\rightarrow 0$ as $h \rightarrow 0$, and C is bounded for $h \leq h_0$.

Substitute (2.11) in the $\sum_{j=1,i} \alpha_{j-1,i} \epsilon_{n-i}$ term of (2.8) to get

$$|\epsilon_N^*| \leq L_h^p(x_{N-1}, h) + (b-a)\Gamma\mu(h) \sum_{i=0}^k |\alpha_{j-1,i}| + hC^* \sum_{n=0}^{N-1} (|\epsilon_n| + |\epsilon_n^*|) \quad (2.12)$$

here C^* is bounded for $h \leq h_0$.

Let $E_N = \text{Max} [|\epsilon_N|, |\epsilon_N^*|]$ and compare (2.11) and (2.12) (noting that $L_h^p(x_{N-1}, h) \rightarrow 0$ as $h \rightarrow 0$) to get

$$|E_N| < h \sum_{n=0}^{N-1} \bar{B} |E_n| + v(h)$$

here \bar{B} is bounded and $v(h) \rightarrow 0$ as $h \rightarrow 0$.

Initially $\epsilon_i = \epsilon_i^* = E_i$ for $i = 0, 1, \dots, k$. Let these be bounded by $K \rightarrow 0$ as $h \rightarrow 0$, then

$$|E_N| < (v(h) + K)(1 + h\bar{B})^N \leq (v(h) + K)e^{(b-a)\bar{B}} \rightarrow 0 \text{ as } h \rightarrow 0 \quad (2.13)$$

Therefore stability and consistency imply convergence.

As in most discussions of convergence, the error bound given by (2.13) is of little practical use. Techniques similar to these in Henrici⁽³⁾ (Section 5.3) can be used to get better asymptotic expressions for the error in particular cases.

3. A Fourth Order Method

If $k = 1$ is used in equations (1.1) and the coefficients are chosen to give the maximum stable order possible, the equations become

$$y_{n+\frac{1}{2}} = y_{n-1} + \frac{h}{8} (9f_n + 3f_{n-1}) + O(h^4)$$

$$y_{n+1}^p = 2y_n - y_{n-1} + \frac{h}{3} (4f_{n+\frac{1}{2}} - 3f_n - f_{n-1}) + O(h^5) \quad (3.1)$$

$$y_{n+1} = y_{n+1}^p - 6\alpha(y_n - y_{n-1}) + \alpha h (f_{n+1} - 4f_{n+\frac{1}{2}} + 7f_n + 2f_{n-1}) + O(h^5)$$

where $0 < \alpha \leq \frac{1}{3}$ for stability.

If it is assumed that $\frac{\partial f}{\partial y} = \lambda$ is constant (if f and y are vectors, λ is a matrix, in which case this analysis must be viewed formally), the equations for the error growth are, in the case of general k ,

$$\epsilon_{n+1}^* = \sum_{i=0}^k (\alpha_{1i} \epsilon_{n-i} + h\lambda\beta_{1i} \epsilon_{n-i}^*) + h\lambda\gamma_{10} \sum_{i=0}^k (\alpha_{0i} \epsilon_{n-i} + h\lambda\beta_{0i} \epsilon_{n-i}^*)$$

$$\epsilon_{n+1} = \sum_{i=0}^k (\alpha_{2i} \epsilon_{n-i} + h\lambda\beta_{2i} \epsilon_{n-i}^*) + h\lambda\gamma_{21} \epsilon_{n+1}^* + h\lambda\gamma_{20} \sum_{i=0}^k (\alpha_{0i} \epsilon_{n-i} + h\lambda\beta_{0i} \epsilon_{n-i}^*)$$

These are a pair of simultaneous difference equations with constant coefficients, each of degree $k+1$. Looking for a solution of the form $\epsilon_n = \xi^n$, $\epsilon_n^* = A\xi^n$, a $(2k+2)$ th degree polynomial equation for ξ is obtained. If $\lambda = 0$, $k+1$ of these roots are 0 and the remainder are the roots of the stability polynomial

$$\xi^{k+1} - \sum_{i=0}^k \alpha_{2i} \xi^{k-i} = 0$$

In the fourth order method given above, these roots are 1 (the principal root) and $1 - 6\alpha$.

This method does not have sufficient accuracy to justify its use in most circumstances, but a simple comparison was made between it with $\alpha = \frac{1}{6}$, the 4th order Runge-Kutta and the 4th order Adams-Bashforth-Adams Moulton predictor-corrector method. The equations

$$y_1 = y_2, y_2 = -y_1, y_3 = y_3, y_4 = -y_4$$

$$y_1(0) = 0, y_2(0) = 1 = y_3(0) = y_4(0)$$

were integrated for $x = 0(.1)50$, and it was found that the method lies between Runge-Kutta and Adams method in accuracy. This is to be expected since Adams used the largest spread of points and Runge-Kutta the smallest, and since the coefficient in the error terms is proportional to the distance of the various points used from the interpolated point. The 5 digit results for $y(50)$ are shown in Table 1.

METHOD	SINX	ERROR	COSX	ERROR	10^{22} EXP(X)	ERROR	10^{-21} EXP(X)	ERROR
Fortran Library	-.26237		.96497		.51847		.19287	
Runge-Kutta	-.26241	- 4	.96495	- 2	.51845	- 2	.19288	+ 1
Hybrid Method	-.26245	- 8	.96494	- 3	.51843	+ 6	.19289	+ 2
Adams	-.26228	+ 9	.96507	+ 10	.51850	+ 3	.19283	+ 4

Table 1. Value at $x = 50$ using points $x = 0(.1)50$ in equations (3.1) calculated on the IBM 7094.

4. A Sixth Order Method

Using equations (1.1) with $k = 2$, and choosing coefficients to get a 5th order method, we arrive at the equations

$$y_{n+\frac{1}{2}} = -\frac{225}{128} y_n + \frac{25}{16} y_{n-1} + \frac{153}{128} y_{n-2} + h \left[\frac{225}{128} f_n + \frac{75}{32} f_{n-1} + \frac{45}{128} f_{n-2} \right]$$

$$y_{n+1}^p = \frac{17}{16} y_n - y_{n-1} + \frac{15}{16} y_{n-2} + h \left[-\frac{11}{16} f_n + \frac{11}{12} f_{n-1} + \frac{5}{16} f_{n-2} + \frac{4}{3} f_{n+\frac{1}{2}} \right] + \alpha_1 z$$

$$y_{n+1} = y_{n+1}^p + (\alpha_2 - \alpha_1)z + \beta W \quad (4.1)$$

where

$$z = -\frac{61}{32} y_n + y_{n-1} + \frac{29}{32} y_{n-2} + h \left[\frac{31}{32} f_n + \frac{41}{24} f_{n-1} + \frac{43}{160} f_{n-2} - \frac{2}{15} f_{n+\frac{1}{2}} \right]$$

$$W = \frac{45}{4} (y_n - y_{n-2}) + h \left[\frac{3}{4} f_n + 16 f_{n-1} + \frac{71}{20} f_{n-2} + \frac{16}{5} f_{n+\frac{1}{2}} - f_{n+1} \right]$$

and $\alpha_1, \alpha_2, \beta$ are constants that must satisfy additional conditions for stability. Each of these equations is 5th order, and the truncation error in y_{n+1} , assuming that y_n, y_{n-1} and y_{n-2} are exact, is

$$\frac{h^6 y^{(6)}}{6!} \left(\frac{29}{4} + \frac{23\alpha_2}{8} - 63\beta \right) + O(h^7) \quad (4.2)$$

Thus along the line $504\beta - 23\alpha_2 = 58$ in the $\alpha_2 - \beta$ plane, the corrector is 6th order.

If we assume that $\lambda = \frac{\partial f}{\partial y}$ is constant, as in Section 2, then the error satisfies a difference equation of order $2k + 2 = 6$. At $\lambda = 0$, 3 of these roots

are 0, the principal root is 1, and remaining two can be seen to satisfy

$$\left(\xi^2 + \frac{61\alpha_2}{32} - \frac{45\beta}{4} - \frac{1}{16} \right) \xi + \left(\frac{29}{32} \alpha_2 - \frac{45\beta}{4} + \frac{15}{16} \right) = 0 \quad (4.3)$$

For the method to be stable, these roots must either lie in the unit circle, or lie on the unit circle and be mutually unequal and unequal to 1. Figure 1 shows the region of stability in the $\alpha_2 - \beta$ plane. Inside the triangle with vertices $A = (-2, +\frac{1}{6})$, $B = (+2, +\frac{7}{45})$ and $C = (+2, +\frac{1}{3})$, the method is stable. The lines AC and BC except for the points A and B correspond to simple roots on the unit circle. The point $D = (1, \frac{59}{360})$ is the point at which both roots of (4.2) are 0, the point of "maximum stability." Also shown in Figure 1 is the line of 6th order accuracy. This crosses the interior of the triangle ABC, meaning that points on the line segment PQ correspond to 6th order stable methods. It is not possible to get a 7th order method by picking a particular point for two reasons. One is that it can be shown not to be stable, even if $f_{n+\frac{1}{2}}$ and f_{n+1} are assumed to be at least of 6th order, the second reason is that $f_{n+\frac{1}{2}}$ can only be of 5th order, and hence $\gamma_{20} h f_{n+\frac{1}{2}}$ will contribute $O(h^7 y^{(6)})$ error.

In choosing a method from the line PQ, one could minimize the largest non-principal root (which then has the value .04). Because many of the preliminary calculations were being done by hand, the point $E = (1, \frac{9}{56})$ which is only a short distance from the minimum, was chosen. This leads to simpler numbers for the human computer. At this point the largest non-principal root is .19.

The choice of α_1 does not affect the results when $\lambda = 0$. However, when $\lambda \neq 0$, α_1 does modify both the error and the stability properties. In order to decide on a suitable α_1 , the 6 roots of the difference equations for y and ϵ_n^* were calculated for various values of α_1 and λh with $\alpha_2 = 1$ and $\epsilon_n = \frac{9}{56}$. The difference between the principal root and the $\exp(\lambda h)$ is a crude measure of the truncation error in one step. $h\lambda_{\min}$, the smallest value of λh for which the largest non-principal root was smaller than the principal root and $h\lambda_{\max}$, the largest value of λh for which the largest non-principal root was less than one, were calculated for various α_1 .

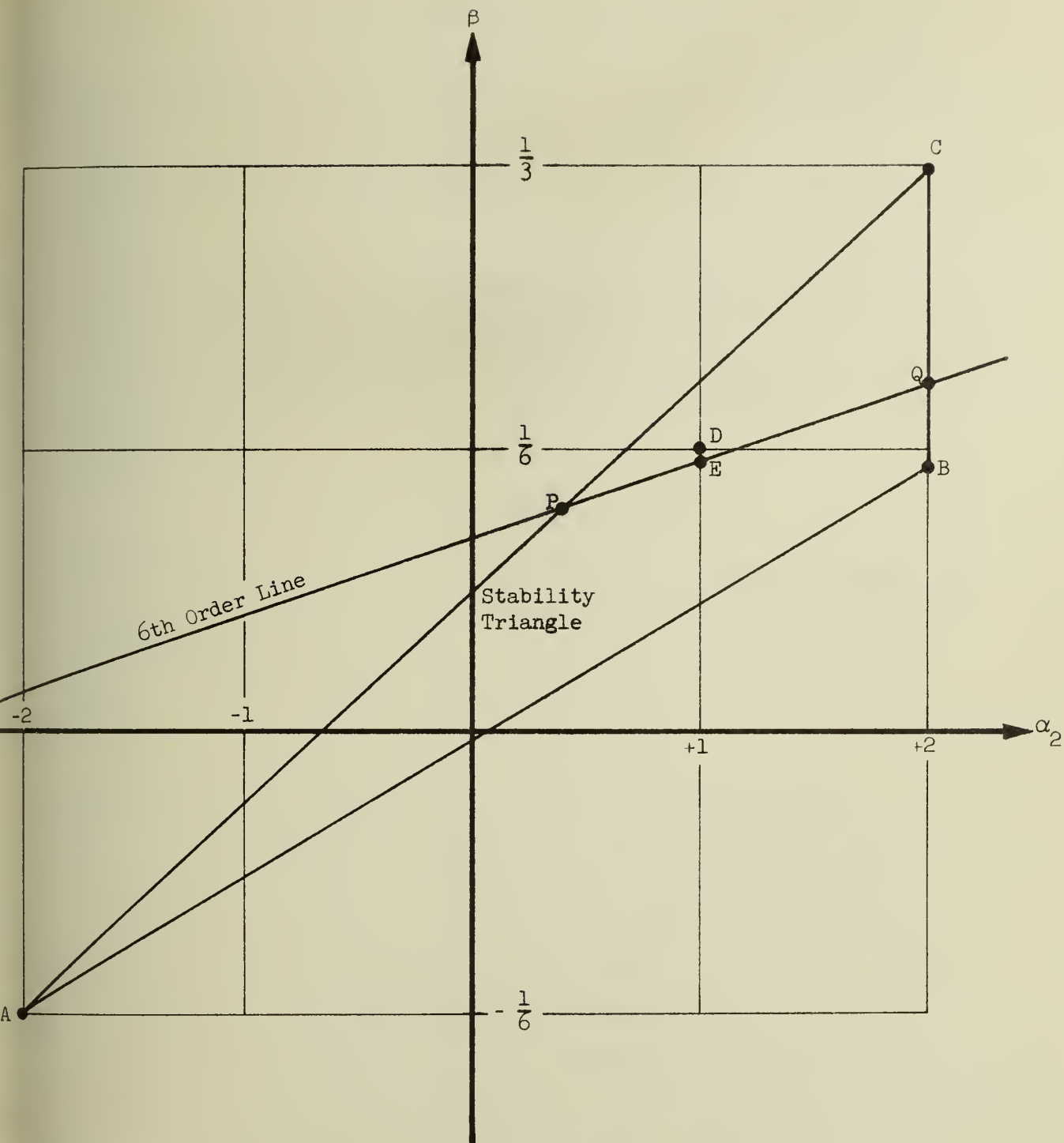


Figure 1. Region of Stability and 6th Order Line.

Also the differences between the principal root and $\exp(\lambda h)$ for $\lambda h = +.1$ and $-.1$, Δ^+ and Δ^- respectively, were calculated for values of α_1 . (Note that at $\alpha_1 = \frac{-58}{23}$, the predictor is also of 6th order, so one expects better accuracy--but not stability--in this neighborhood.)

The results are shown in Figures 2 and 3. As was expected, the error decreases as α_1 moves down towards $-\frac{58}{23}$, but the range of λh for which the equation is stable decreases. Any α_1 between $-.5$ and $+.4$ would appear to be a reasonable choice, the former being preferred if the equation is unstable, the latter for stable equations.

In order to compare different parameter combinations for $k = 2$, the equations

$$y_1 = y_2, y_2 = -y_1; y_1(0) = 0, y_2(0) = 1$$

were integrated for $x = 0(.1) 200$ on an IBM 709⁴. Four parameter combinations were used, corresponding to the points D (5th order, "maximally stable") and E (a 6th order method) with $\alpha_1 = 0$ and 1 in each case. The results are shown in Table 2. As expected, the 6th order method using $\alpha_1 = 0$ is slightly superior.

Nordsieck⁽⁴⁾ gives the results of integrating $J_{16}(X)$ from $x = 6$ to $x = 6138$ by his method which is a variant of the Adams 6th order process. Therefore, this equation has been integrated from the same starting values that Nordsieck used. ($J_{16}(6) = .000\ 001\ 201\ 950$; $dJ_{16}^{(6)}/dx = .000\ 002\ 986\ 480$.) The integration was carried out by Runge-Kutta, Adams-Bashforth-Adams-Moulton 6th order (with single correction where it was stable ($h = \frac{1}{32}$) and double correction where single correction was unstable ($h = \frac{1}{16}$ and $h = \frac{1}{8}$)), by the 6th order hybrid with $k = 2$ corresponding to the point E in Figure 1 with $\alpha_1 = 0$, and by an 8th order method discussed in the next section. The step sizes $1/8$, $1/16$ and $1/32$ were used in order to make meaningful comparisons with Nordsieck's method. The results that he quotes, although for variable step size, took average step sizes of $1/8$ and $1/16$. Since this method involves two evaluations of the derivative, it seemed pertinent to use Adams single corrector with one derivative evaluation for $h = 1/32$. All of the methods (except for R-K) were started by an R-K integration on $1/8$ the step size in order not to introduce starting errors into the comparison. The results are summarized in Tables 3 and 4.

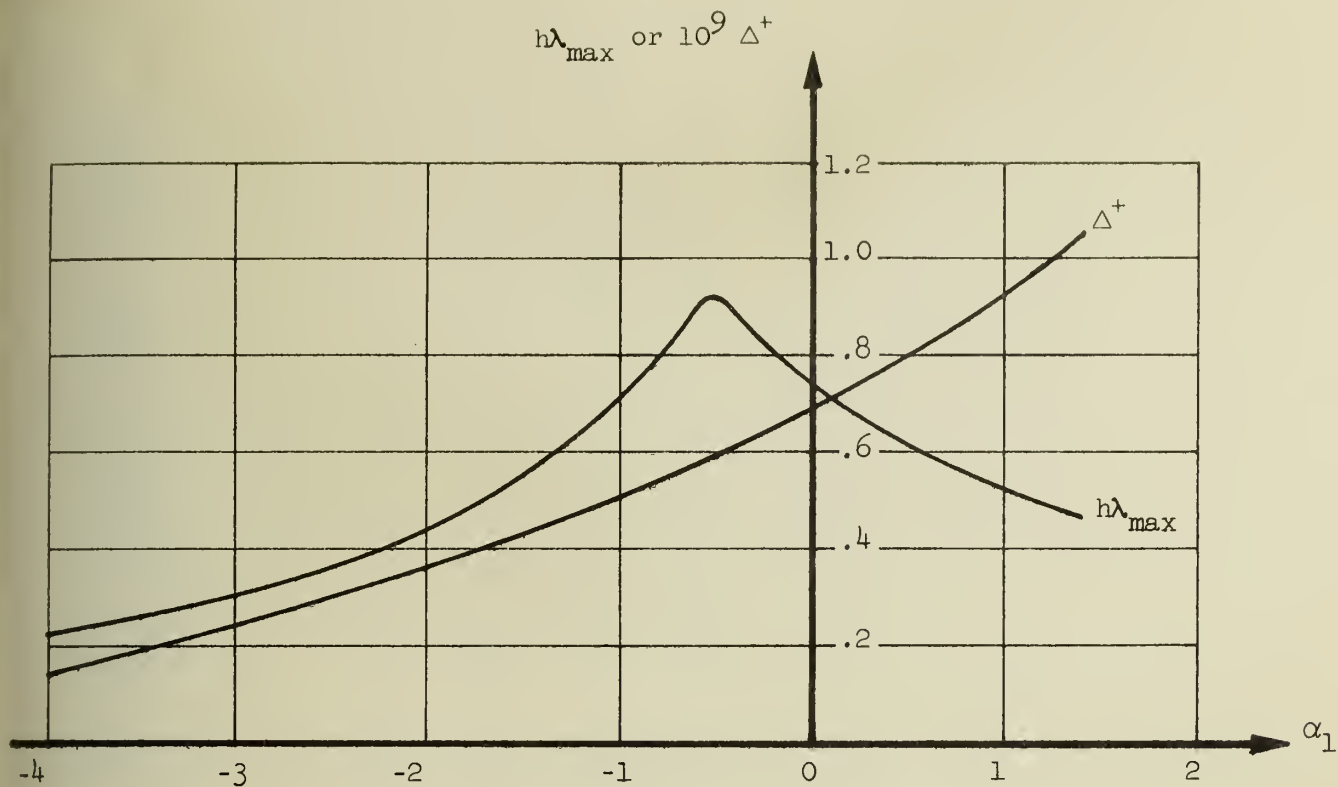


Figure 2. Δ^+ and $h\lambda_{\max}$ against α_1 .

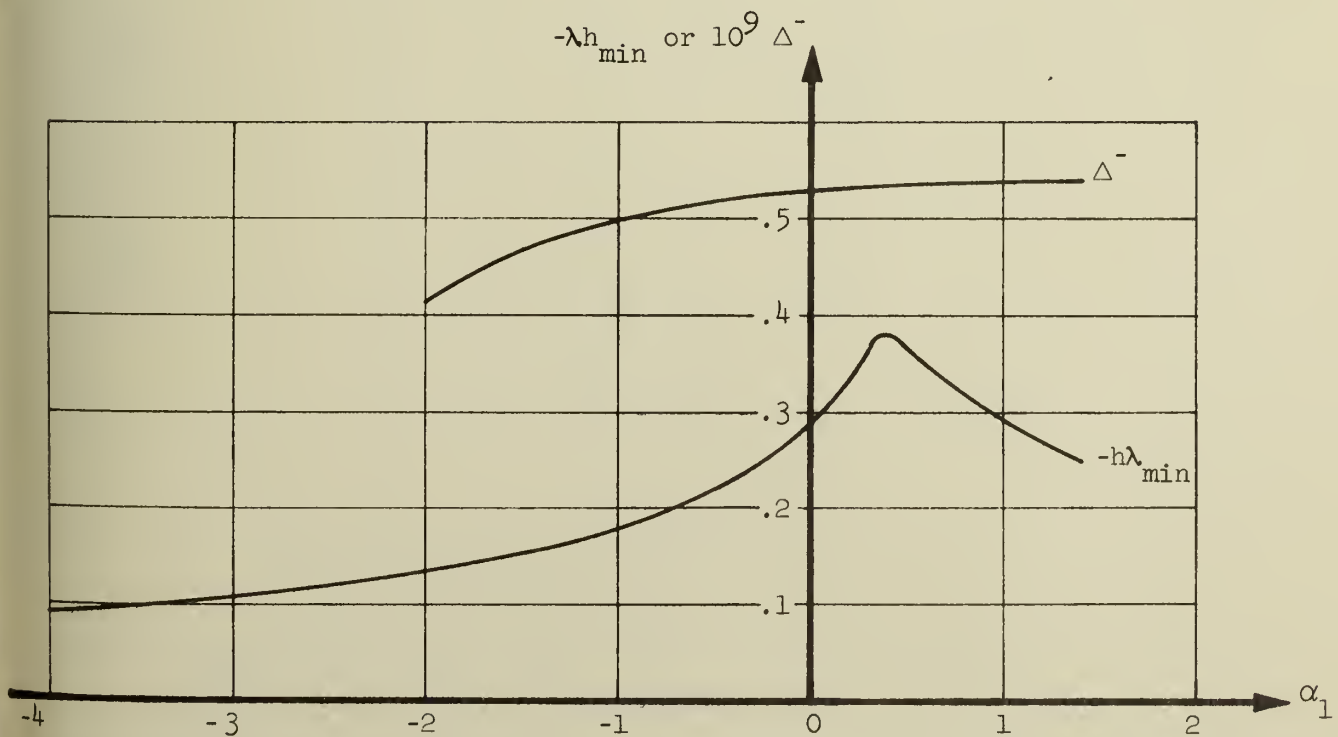


Figure 3. Δ^- and $-h\lambda_{\min}$ against α_1 .

		x = 100			x = 200		
		SIN	ERROR	COS	ERROR	SIN	ERROR
"TRUE" (IBM 7094 Fortran)		-5063 656		8623 186		-8732 973	4871 871
6th Order (E) $\gamma_1 = 1$		506	150	051	135	575	398 795
6th Order (E) $\gamma_1 = 0$		557	99	059	127	671	302 757
5th Order (D) $\gamma_1 = 1$		489	167	042	144	528	445 794
5th Order (D) $\gamma_1 = 0$		543	113	052	134	641	332 760
							111

Table 2. Comparison of k = 2 methods x = 0 (.1) 200.

Step Size h	ADAMS*	6th Order Hybrid	NORDSIECK	RUNGE-KUTTA	8th Order Hybrid
1/32 (ERROR)	-974 5831.7	-974 5831.8		-974 5671.7	
	-1	-1		159	
1/16 (ERROR)	-974 5809.0	-974 5839.0	-974 5792	-974 3089.3	-974 5831.0
	22	-8	39	2742	0
1/8 (ERROR)	-974 3394.7	-974 6398.0	-974 1657	-969 5556.4	Unstable
	2436	-567	4174	5 0275	

* Single correction for h = 1/32. Double correction for h = 1/16 and 1/8.

Table 3. $10^9 J_{16}$ (6136) by integration of $y + y/x - \left(\frac{x^2}{256} - 1 \right) y = 0$ for $x = 6(h)$ 6136 .

Step Size h	ADAMS*	6th Order Hybrid	NORDSIECK	RUNGE KUTTA	8th Order Hybrid
1/32 (ERROR)	136 2486.3 1	136 2485.5- 0		136 1994.4 -491	
1/16 (ERROR)	136 2475.6 -9	136.2509.9 +25	136 2434 -51	135 4615.3 -7870	136 2485.4 0
1/8 (ERROR)	136 2409.8 -75	136 4177.6 1693	135 9819 -2666	123 6316.9 12 6068	Unstable

* Adams was single corrector h = 1/32. Double corrector h = 1/16 and 1/8.

Table 4. $10^9 J_{16}$ (6138) by integration of $y + y/x - \left(\frac{x^2}{256} - 1 \right) y = 0$ for $x = 6(h)$ 6138.

The values of J_{16} (6138) and J_{16} (6136) quoted by Nordsieck are $10^{-9} \times 1362485$ and $-10^{-9} \times 9745831$ respectively. The errors shown in the tables are based on these values and the calculated values rounded to 7 significant digits and thus can be in error by ± 1 . The results were obtained by single precision floating point arithmetic on ILLIAC II, which represents about 13 decimal digits (22 base 4 digits). Therefore the roundoff in the integration should not be significant in the comparisons.

5. Higher Order Methods

For $k = 1$ and $k = 2$ it was possible to achieve $(2k + 2)$ th order accuracy with equations (1.1). This relation can also be seen to hold for $k = 0$ when the equations represent the second order Runge-Kutta method. The question of whether this is true for all k naturally arises. Unfortunately the answer is no, but it is true for $k = 3$ and 4 , giving rise to 8th and 10th order methods respectively. Parameters for both cases are given below, but the 10th order method has undesirable stability properties due to large roots of the stability polynomial and hence it is not recommended. For $k = 5, 6, 7, 8$, and 9 , methods of order $2k + 2$ are not stable.

These results were obtained by a numerical investigation of methods for $k = 3, 4, \dots, 9$ on ILLIAC II. For each k , the coefficients of the method were calculated from the $2k + 3$ linear simultaneous equations obtained by a Taylor series expansion to $2k + 3$ terms. Since there are $2k + 4$ parameters in the corrector, these coefficients are linear functions of one parameter, which was taken as γ_{21} , the coefficient of f_{n+1} . The roots of the stability polynomial

$$\xi^{k+1} - \sum_{i=0}^k \gamma_{20}(\gamma_{21}) \xi^{k-i} = 0$$

were then studied.

If $\alpha_{20}(\gamma_{21})$ goes to ∞ , as $|\gamma_{21}|$ goes to ∞ , the largest root is asymptotic to $\alpha_{20}(\gamma_{21})$. The remaining roots are $O(1)$ as γ_{21} becomes infinite. Therefore, it is only necessary to examine values of γ_{21} in a region near the origin to try to find stable methods. The searching method was crude, consisting of a series of runs, each calculating the roots of the stability polynomial for a regular mesh of points on the γ_{21} axis. The first run took a fairly large interval between mesh points in order to locate the region where the dominant

root would be small; successive mesh refinements narrowed in on this area until it was possible to plot the roots of the equation. When k is larger than 4, the largest root of the equation fails to get below +1 as γ_{21} is increased from $-\infty$ before a second root of the equation exceeds +1. These two roots coalesce and become complex conjugates, remaining outside of the unit circle, then coalesce again on the negative real axis and become real, one of them then goes off towards $-\infty$.

Particular cases of the method for $k = 3$ and 4 are given in Table 5. The largest non-principal roots of these methods are .274 and .695 respectively. The latter is so large as to make that method unstable for values of λh much different from 0, so it has not been used in any further work. The 8th order method was used to integrate $J_{16}(x)$. The results at $x = 6136$ and $x = 6138$ are shown in Tables 3 and 4 for a step size of $1/16$. The method is unstable for $h = 1/8$, and $h = 1/32$ was not tried because of the accuracy at $h = 1/16$. As in the case of $k = 2$, there is a degree of freedom in the predictor which is not shown in Table 5. It is possible that the 8th and 10th order methods could be made more stable for some values of λh by a better choice of the predictor. For example, a multiple, α_1 , of

$hf_{n+\frac{1}{2}}$	+ 28.3311	6319 ⁴	y_n
	+ 17.6367	1875	y_{n-1}
	- 37.3242	1875	y_{n-2}
	- 8.6436	6319 ⁴	y_{n-3}
	- 11.2565	1041 ⁶	hf_n
	- 43.3398	4375	hf_{n-1}
	- 27.1523	4375	hf_{n-2}
	- 2.1940	1041 ⁶	hf_{n-3}

may be added to the $k = 3$ predictor without changing the order from 7. The error term of the predictor is proportional to

	Half Point (j=0)	Predictor (j=1)	Corrector (j=2)	
α_{j0}	-3.987630208 $\dot{3}$	-.28030 $\dot{3}$	1.3450454 $\dot{5}$	} k=3 8th Order
β_{j0}	2.392578125	-.84090 $\dot{9}$	-.0976636 $\dot{3}$	
α_{j1}	-2.392578125	-9.61363 $\dot{6}$	-.3129545 $\dot{4}$	
β_{j1}	7.177734375	7.15909 $\dot{0}$	-.1936636 $\dot{3}$	
α_{j2}	6.029296875	8.15909 $\dot{0}$	-.0468636 $\dot{3}$	
β_{j2}	4.306640625	7.37727 $\dot{2}$.0136090 $\dot{9}$	
α_{j3}	1.350911458 $\dot{3}$	2.73484 $\dot{8}$.0147727 $\dot{2}$	
β_{j4}	.341796875	.7175324 $\dot{6}$.0041415584 $\dot{4}$	
γ_{j0}	-----	1.4961038	.76301298 $\dot{7}$	
γ_{j1}	-----	-----	.148200000	
α_{j0}	-6.5608 97827	-2.2091 62227	1.8125 58911	} k=4 10th Order (Rounded to 9 digits)
β_{j0}	3.0281 06689	-1.1450 26643	-.3513 5364	
α_{j1}	-16.1499 02344	-45.7510 95327	-.5903 09059	
β_{j1}	16.1499 02344	24.4532 85971	-.6104 06157	
α_{j2}	8.7209	7.1545 29309	-.4139 11190	
β_{j2}	21.8023 68165	53.2631 61639	.0758 11013	
α_{j3}	13.5131 83594	37.0961 51572	.1644 99704	
β_{j3}	6.9213 86719	20.3444 81098	.1059 25400	
α_{j4}	1.4766 69312	4.7095 76673	.0271 61634	
β_{j4}	.3364 56299	1.0920 36708	.0063 09216	
γ_{j0}	-----	1.6767 85926	.8161 28202	
γ_{j1}	-----	-----	.1416 00000	

Table 5. Stable Methods for k = 3 and 4.

in large vector problems, then the 8th order hybrid method is more accurate but still requires fewer additional points. If one is prepared to use more function evaluations and/or use other points besides the half point, probably higher order stable methods exist, and would pay off in special cases, but it seems unlikely that one would want to go beyond order 8 or 10, as it appears to become harder to control the larger number of extraneous roots, or to have much knowledge of the high order derivatives.

Gragg and Stetter⁽²⁾ consider correctors similar to the corrector used in this paper with $J = 2$. They choose the point of evaluation of the first predictor to optimize the (stable) order of the corrector. This introduces an additional degree of freedom into the corrector, which, with the degree of freedom already existing in the correctors used here allows a corrector of order $2k + 4$ rather than $2k + 2$ to be chosen. Their surprising result is that for $k \leq 3$ (the k of Gragg and Stetter is one larger than this one), these optimal order methods are stable. The result for $k \geq$ is not yet known. Because the new corrector order is greater, the predictors must be of a correspondingly higher order, meaning that additional mesh points must be used. Since these additional points are needed, it would seem desirable to look for methods which make use of them to reduce the error term or to put the additional non mesh point, or points, at a desirable location.

$$\frac{d^8 y}{dx^8} h^8 \left[576 + 288.75 (\alpha_1 + 1.49 \ddot{6} \ddot{1} \ddot{0} \ddot{3} \ddot{8}) \right]$$

Positive α_1 would improve the stability slightly since it tends to increase α_{10} and α_{11} from negative values, and to decrease the other α_{1i} from positive values. However, positive α_1 will increase the predictor error, and thus a large error will result if $h\lambda$ is not zero.

6. Conclusion

The most obvious conclusion that can be drawn from Tables 3 and 4 is that one should not use Runge Kutta if high order derivatives are well behaved! For the general problem, it is difficult to say much more. If the evaluation of the derivative is time consuming, then it seems reasonable to compare Adams' single corrector $h = \frac{1}{32}$, and the other methods (including Adams double corrector) for $h = \frac{1}{16}$. Each of these methods takes 32 evaluations per unit step in x . Nordsieck's method differs from Adams's method only in the predictor and in the effect when the step size is changed, but this apparently introduces a considerable amount of error. Automatic step size changing probably would pay off in cases where there is a sudden change of behaviour in the function; for those cases one expects Nordsieck's method to be superior. A step changing mechanism can be programmed on the basis of the difference between the predictor and corrector of the hybrid methods presented. Although this is a term of order $h^{(2k+2)} y^{(2k+2)} + O(h^{2k+3})$, it is indicative of the behaviour of the "average" value of the error term which includes terms in $h^{2k+3} y^{(2k+3)}(x)$ and $h^{2k+3} \lambda y^{(2k+2)}(x)$.

The hybrid 6th order has about three times the error of Adams double corrector in one case, and about 1/3 in the other. It has the advantage of using two additional points instead of 4 (5 if Adams-Bashforth is used as a corrector). If storage space and hence the number of points is a criterion

BIBLIOGRAPHY

1. Fox, L. Numerical Solution of Ordinary and Partial Differential Equations. Pergamamon Press. London, (1962).
2. Gragg, W. B., and Stetter, H. J. Generalized Multistep Predictor Corrector Methods. JACM. 11, 2. (April, 1964) pp. 188-209.
3. Henrici, P. Discrete Variable Methods in Ordinary Differential Equations. Wiley, New York. (1962).
4. Nordsieck, A. On Numerical Integration of Ordinary Differential Equations. Math. Comp. 16, 7. pp. 22-49, (Jan. 1962).



UNIVERSITY OF ILLINOIS-URBANA
510.84 IL6R no. C002 no.160-170(1964
Report /



3 0112 088398166